A Case Study in Extracting DEMs from High-Resolution Mars Stereo Pairs Using a Simple Computer Vision Algorithm

Ming-Li Lin (mingli@cs.brown.edu)
Department of Computer Science, Brown University

Abstract

Background: Geoscientists have been doing research on Mars for decades. As the photography technique has improved, we now have thousands of high-resolution stereo pair images but it is difficult to extract quantitative information including relative heights and slopes without the aid of computer vision. Our goal is to create an easy-to-use, accurate and high performance tool which can process a medium size(6000x6000) Mars stereo pair on a standard desktop, and transforms stereo pairs into digital elevation maps for geoscientists.

Method: We applied a simple computer vision algorithm which could be implemented in C++ in 1500 lines to compute elevation maps from stereo pairs. The computer vision algorithm produced a disparity map and we used position data from the satellite that produced the stereo pair to create a digital elevation map with height measures in meters.

Conclusion: We ran three tests, Victoria Crater, Centauri, and Gale Crater stereo pairs. The details of the stereo pairs are listed below. The accuracy values are the comparisons between our results and the ground-truth data provided by NASA[3].

Name Detail	Victoria Crater	Centauri	Gale Crater
Image Size	2000x2000	6824x6236	14011x13099
Maximum Height Difference (computed with SOCET Set software)	79m	477m	506m
Computational Time Using Our Tool	45min	22hours	~5days
Our Algorithm's Accuracy compared with SOCET Set result (measured by differences less than 15 meters)	94.97%	89.11%	87.46%

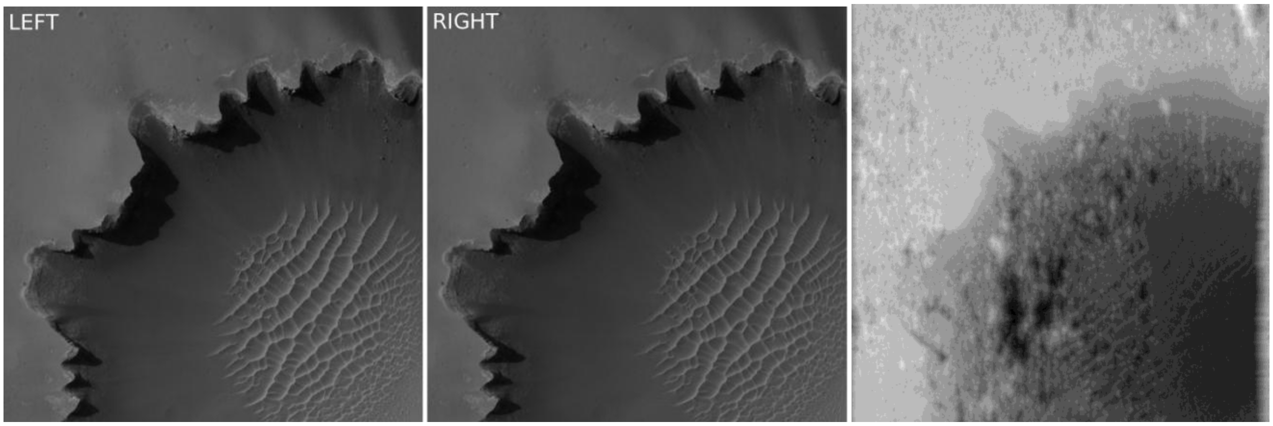

Figure 1. Our challenge was to make it easier to convert a stereo pair (left two images) into an accurate digital elevation map presents as a image. (rightmost image).

I. Introduction

Doing quantitative analysis using stereo pairs is important but still a challenge for geoscientists.

Because the images of the stereo pairs on Mars were not taken at the same time, the shadow and terrain changed due to the time differences, which cause the stereo pair to contain noise. The alignment of the stereo pair is also difficult because of the specialty of the cameras in the satellite.

Each image of the Mars stereo pairs is not taken by only one camera. There are several cameras which produce images as long strips. We need to combine these long strips to one single image and then do the projection to the ground. It causes imperfect alignments which makes our project even more difficult than a typical stereo pair problem.

The standard procedure in computer vision to deal with the stereo pair problem is: (1) Image Rectification (alignment) (2) Correspondence Problem (Disparity Calculation) (3) Depth Reconstruction. The algorithm in our tool is based on Sun's work[2]. It takes the processed but imperfect stereo pair and creates the depth map in meters.

The following sections present how our tool works. Section II and III present why the algorithm we adopted is good for the stereo pairs on Mars. Section V and VI present our results and discuss what we observed in the results.

II. Related Work

A significant amount of research has been done on creating DEM's from stereo pairs, but still there is no inexpensive, easy-to-use tool available for geoscientists to use. SOCET SET [3] is the state-of-the-art commercial tool used by NASA scientists for creating DEMs from stereo pairs, but it is expensive ($100k) and can take weeks to process one image. A fast, inexpensive and widely-available tool would have a significant impact on geosciences research.

We considered implementing several published methods. The correspondence problem could be solved by minimizing an objective function which is composed by smooth constraint, intensity correlation component and shading factor. In Fua's work, not only the occlusion could be solved during the reconstruction of the 3D mesh by a hidden surface algorithm, but also the albedo of the surface doesn't need to be constant[1].

The approach could get a near optimal solution and create the 3D model directly. However, the computational complexity relates to the number of pixels per image and number of facets and number of samples per facets. The stereo images from Mars could be very large (more than 15,000x15,000 pixels for one image) and the terrain could be very complicated. Therefore, it is not an efficient way to deal with the Mars stereo pairs.

A lot of work has been proposed in computer vision area, and we need a method which memory requirement and computational time is linear to the image size and disparity value.

III. Solving Correspondence Using Computer Vision

Input Assumption:

Although the alignment of rows in Mars stereo pair images is not always in perfect correspondence, we still use the assumption that the input is perfectly aligned and ignore the alignment errors. We will discuss the consequence of the alignment errors in section VI.

The following presents the basic idea of the Sun algorithm [2] that we based our work on. First, Figure 2 is the illustration of the alignment. If the stereo pair are perfectly aligned, a correspondence point of the point (a, b) in the left image must lie on the line y=b in the right image.

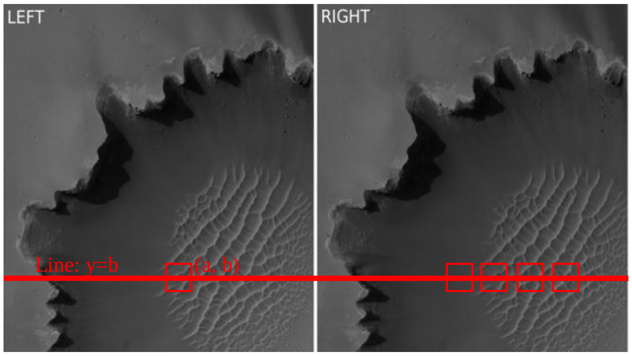

Figure 2. Illustration of the image alignment.

Process of Our Computer Vision Method:

The complexity of our algorithm for both time and memory is O(MND') where MN is the size of images and D' is a value much less than the actual disparity range D in the stereo pair.

The basic idea of the algorithm is simple:

1. Allocate a 3D matrix MxNxD and calculate the similarity values of the stereo pair for each pixel and its disparity values.

   For example, if we know the correspondence point of a point (a, b) on the left image must lie in the range of (a-d ~ a+d', b) on the right image, the disparity range D is (d'+d), and we can calculate a similarity value for each possible disparity of the pixel. Therefore, for each pixel, we can have D similarity values, and that's why we need a 3D matrix.

2. Fill in the 3D matrix with similarity values and find the 3D surface which could contain the maximum accumulated similarity value.

3. The disparity values of pixels in the image would

correspond to the D value that the surface contains in the memory. In our Mars cases, the surface would be a scale to the terrain. The reason is explained in section VI.

The following picture shows the possible result of a maximum-surface.

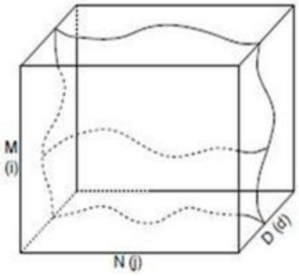

Figure 3. The illustration of the maximum-surface.[2]

Similarity Measurement:

There are many approaches to calculate the similarity. We used cross correlation coefficient as the similarity measurement. It is more popular than SAD (sum of absolute differences) and SSD (sum of square differences) because it corresponds to optimal signal-to-noise ratio estimation [2].

The zero mean normalized cross correlation value on the point (i, j) with the disparity d for the right image can be written as follows:

$$\mathrm{C(i,j,d)} = \frac{\mathrm{cov}_{\mathrm{i,j,d}}(\mathrm{f,g})}{\sqrt{\mathrm{var}_{\mathrm{i,j}}(\mathrm{f})} \times \sqrt{\mathrm{var}_{\mathrm{i,j,d}}(\mathrm{g})}}$$

where f and g are the intensity values of the left and right images, which size are both M x N. The equations of the covariance and variance used are as follows:

$$\mathrm{cov}_{\mathrm{i,j,d}}(\mathrm{f,g}) = \sum_{\mathrm{m=i-K}}^{\mathrm{i+K}} \sum_{\mathrm{n=j-L}}^{\mathrm{j+L}} \left(\mathrm{f}_{\mathrm{m,n}} - \bar{\mathrm{f}}_{\mathrm{i,j}}\right)\left(\mathrm{g}_{\mathrm{m-d,n}} - \bar{\mathrm{g}}_{\mathrm{i-d,j}}\right)$$

$$\mathrm{var}_{\mathrm{i,j}}(\mathrm{f}) = \sum_{\mathrm{m=i-K}}^{\mathrm{i+K}} \sum_{\mathrm{n=j-L}}^{\mathrm{j+L}} \left(\mathrm{f}_{\mathrm{m,n}} - \bar{\mathrm{f}}_{\mathrm{i,j}}\right)$$

$$\mathrm{var}_{\mathrm{i,j,d}}(\mathrm{g}) = \sum_{\mathrm{m=i-K}}^{\mathrm{i+K}} \sum_{\mathrm{n=j-L}}^{\mathrm{j+L}} \left(\mathrm{g}_{\mathrm{m-d,n}} - \bar{\mathrm{g}}_{\mathrm{i-d,j}}\right)$$

where L and K define the correlation window size, and $\bar{\mathrm{f}}, \bar{\mathrm{g}}$ are the mean value within the local windows. In order to fully utilize the box filter technique, we rewrite the covariance equation to:

$$\mathrm{cov}_{\mathrm{i,j,d}}(\mathrm{f,g}) = \sum_{\mathrm{m=i-K}}^{\mathrm{i+K}} \sum_{\mathrm{n=j-L}}^{\mathrm{j+L}} \left(\mathrm{f}_{\mathrm{m,n}} \times \mathrm{g}_{\mathrm{m-d,n}}\right) - \mathrm{W}\left(\bar{\mathrm{f}}_{\mathrm{i,j}} \times \bar{\mathrm{g}}_{\mathrm{i-d,j}}\right) \text{ [2]}$$

where $W = (2\mathrm{K} - 1)(2\mathrm{L} - 1)$

The similarity value we get would be always between -1~1.

Advantage of Pyramid and Our Memory Carving Technique:

Pyramid technique is broadly used in computer vision. The idea is to exploit the information from coarse data and use it in calculating the finer data. It is also very useful for the method used in our tool.

Figure 3 is the illustration of the memory we need without any memory carving. Sun used a technique called "Rectangular Subregioning Process" to divide the 3D memory into subregions and only calculated the similarity values in the subregions[2]. However, it's hard to know if the region we selected is good enough and the approach also doesn't fully utilize the advantage of the pyramid.

Our approach of memory carving is to use the disparity map from the coarse data as the prediction for the finer data. Figure 4 is a 2D example of our approach. The first image is the coarse data. The second image is the scaled up coarse data and we could carve the memory space around the line to calculate the finer line at the current pyramid level by using the carved memory, which is shown as the third image.

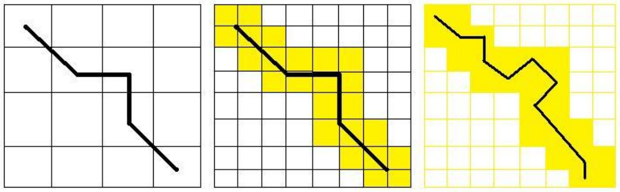

Figure 4. A 2D example of our memory carving approach. The yellow part is the memory we used in our method.

The idea for the 3D memory is the same. We used the disparity map calculated from the last level of the pyramid and carved the memory around the maximum-surface to do the calculation at the current level.

By using this approach, the memory space we used would also be a surface which surrounds the coarse disparity map. Therefore, the memory we carved would be more flexible than Sun's approach.

Another advantage of the pyramid is the temporary results produced during the process. Theoretically speaking, the final result of the method should be the most accurate one if the stereo pair images are perfectly aligned and noiseless. However, due to the imperfect stereo pairs of Mars, the stereo pair contained noise especially from the alignment problem. The scaled stereo pairs are smaller but there are also fewer errors. Therefore, the results of the scaled data would also show less error. The Gale Crater is a good example and will be discussed in section VI.

Algorithm for Solving Maximum-Surface:

Sun claims that he developed an algorithm called TSDP: two-stage dynamic programming, which provides an optimal solution for obtaining the disparity map from the 3D volume. The first stage of the algorithm is to obtain an accumulated similarity 3D volume in the vertical direction for each vertical slice and the second stage is to create the maximum-surface by using the 3D volume obtained from the first stage [2]. It is an efficient method and we adopted the algorithm in our tool.

IV. Use satellite information

Because the satellite is very high to the ground (hundreds of kilometers), and the terrain in an image is relatively small (few kilometers at most), we could assume the light beams are parallel to each other.

With the parallel assumption, the disparity-to-meter problem is easy to solve. The emission angles of the satellites are known and the meters per pixel in the image is also known. The way in which disparity transfers to meters is just a simple geometry problem.

Figure 5 represents the geometry problem. x and y are the emission angles. The equation would be:

$$\begin{aligned} &|\tan(x)+\tan(y)| * height\ per\ disparity \\ &= meters\ per\ pixel\ on\ the\ image \end{aligned}$$

where x and y are the emission angles.

If the target is not in the middle of two satellites, we could just replace the addition between the tangents to subtraction, and the problem could still be solved easily.

Since the disparity map is the scale to the depth map because the light beams are parallel, the maximum-surface founded would be the scale of the terrain.

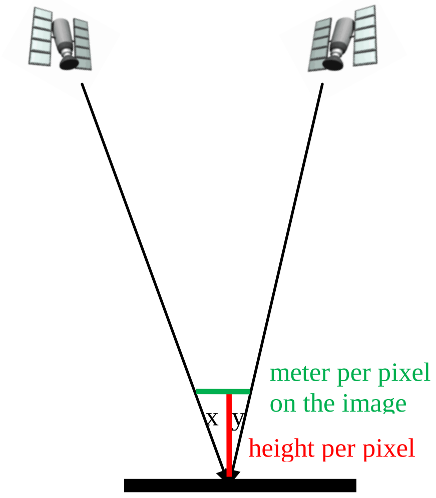

Figure 5. The illustration of the satellites.

V. Results

Our software implementation was done in C++ and consists of 1500 lines of code. The inputs to the algorithm are the two stereo pairs and the satellite information.

We used three well-studied locations on Mars to evaluate our method: Victoria Crater, Centauri, and Gale Crater on Mars. Our evaluation method was to compare the meter values we computed with those calculated by SOCET Set.

We did not take the missing pixels (i.e., pixels that appeared in either the left or right image but not the other) into consideration in our comparisons. The missing pixels problem is discussed in section VI.

Table 1 is the statistical data which show the differences between our results and the ground-truth provided by NASA. Our tool created pretty good results for the geologists as shown in the table.

Name Difference	Victoria Crater	Centauri	Gale Crater
<1m	0.1778	0.0805	0.0855
<2m	0.3467	0.1608	0.1674
<3m	0.5009	0.2413	0.2474
<4m	0.6312	0.3201	0.3309
<5m	0.7250	0.3957	0.4159
<6m	0.7860	0.4673	0.4945
<7m	0.8280	0.5347	0.5630
<8m	0.8585	0.5969	0.6220
<9m	0.8824	0.6538	0.6733
<10m	0.9004	0.7054	0.7184
<11m	0.9148	0.7519	0.7571
<12m	0.9263	0.7938	0.7896
<13m	0.9355	0.8317	0.8193
<14m	0.9432	0.8640	0.8481
<15m	0.9497	0.8911	0.8746

Table 1. The coverage percentages comparing to the ground-truth provided by NASA.

The top-left image in Figure 6 is the original image of Centauri. The top-right image shows the difference between our result and the ground-truth. The white parts represent the differences larger than 15 meters. The missing pixels we ruled out spread on the far left and the far right of the image. The bottom-left image is the ground-truth and the bottom-right image is our result.

Our tool created a very similar result to the ground-truth from NASA as shown in Figure 6.

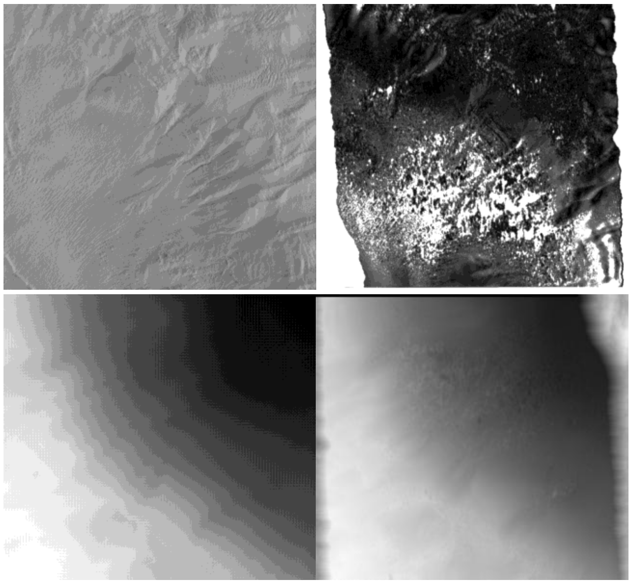

Figure 6. The results of Centauri on Mars. Our tool can create a very similar result to the ground-truth provided by NASA. The bottom two images are the results from NASA (Left image) and our tool (Right image).

VI. Discussion

Missing Pixels Problem:

The missing pixels are the pixels that cannot be matched to any pixels at another image of the stereo pair because the matching pixels are out of the image. It is a problem that any stereo pair would have. However, the missing pixels usually cover several percent of the image, and the TSDP algorithm would be largely affected by the missing pixels and get a lot of error data on the edge of the image as shown in Figure 6.

We believe our results could be improved significantly if the results from this method could be combined with an interactive system where geoscientists could label disparity values manually on the image. Once the manual data could be recognized by the program, we could solve the missing pixels problem.

Imperfect Projection/Alignment and Similarity Values:

Perfect alignment is our main assumption for the input. However, none of the stereo pairs we used in this paper is perfectly matched.

Our tool produced good results for Victoria Crater and Centauri but it didn't actually create a good depth map for Gale Crater, although the statistic data seem to be accurate.

The top image in Figure 7 is the depth map created by our tool. There are weird strips on the image which are very obvious if we look at the difference image and the similarity image at the bottom of Figure 7.

The bottom-left image in Figure 7 is the same as the top-right image in Figure 6. The white parts represent the differences larger than 15 meters. Because the Mars stereo pairs are combined by many strips, it's difficult to align them for a large image like Gale Crater. The imperfect aligned pixels caused the calculation error in our program and created the weird strips in the depth map.

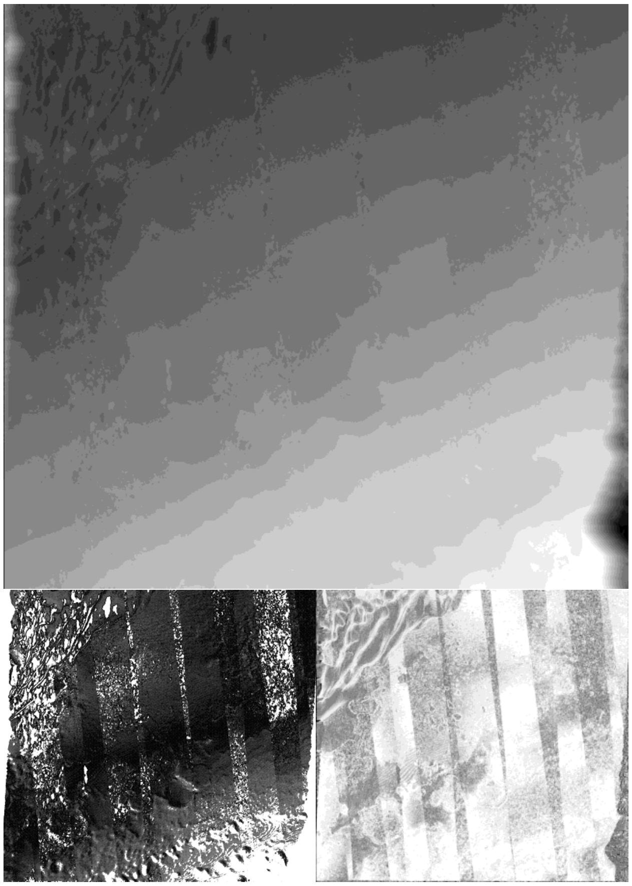

Figure 7. The errors caused by imperfect alignment.

Our tool can handle large images but cannot deal with the huge alignment errors at this stage. The error might be solved by the coarse results of the pyramid since the phenomenon is imperceptible for coarse data as shown in Figure 8. However we haven't developed an effective approach.

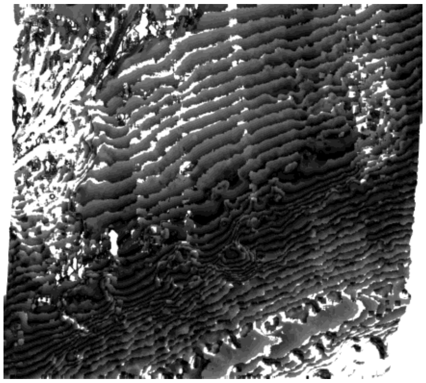

Figure 8. The imperfect alignment effect is slight for coarse data.

After the disparity map was created by our program, we could create a similarity matrix because each value in the disparity map matches to a similarity value in the 3D memory we mentioned in section III. We can create the similarity image (shown in Figure 7) by the similarity matrix and use the similarity image to know which parts in the image our tool has less confidence in.

The similarity image is a useful reference for the confidence level in the result and it is very sensitive to missing pixels and imperfect alignments. Geoscientists could use it to rule out the incorrect parts of the results due to these two factors.

VII. Conclusion

We implemented and tested a fast stereo matching computer vision algorithm to produce digital elevation maps from stereo pairs. The algorithm produces a disparity map which we transform into a digital elevation map that represents relative heights in meters. This enables quantitative analyses such as measuring relative heights or reporting slopes. The algorithm is accurate for the stereo pairs with small alignment errors and also provides similarity images as a tool for the confidence in the results.

We believe that our tool would be very convenient for geoscientists.

VIII. References

[1] P. Fua, Y. G. Leclerc. (1995). *Object-centered surface reconstruction: Combining multi-image stereo and shading.* International Journal of Computer Vision, 16, 35-56

[2] Changming Sun. (2002). *Fast Stereo Matching Using Rectangular Subregioning and 3D Maximum-Surface Techniques.* International Journal of Computer Vision, vol47, pp. 99-117.

[3] http://www.socetgxp.com/content/products/socet-set
SOCET Set is a commercial software that NASA used to create ground-truth data.

Possible Future Extensions:

(1) Solve the missing pixels problem:

We applied the TSDP method in Sun's work so we got the same missing pixels problem as the results in Sun's paper. However, he didn't mention it in the paper and the problem is hard to solve.

The missing pixels problem is caused by empty similarity values. We need to give similarity values to the missing pixels first so we can use TSDP without the problem.

The question is: "How to assign similarity values when the corresponding pixels are out of the image?" We tried to copy the values on the edge of the image to the missing pixels but the results were not good. However, it is an initial idea for solving the problem.

(2) Imperfect projection/alignment problem:

We need a better projection program to do the alignment. Although we provide the "coarse result" idea for solving the problem, there is no better way than fixing the projection itself.

There are many papers in computer vision discussing the projection methods. However, we haven't found one paper which takes the multi-camera issue into consideration. The main issue for Mars stereo pair is that one image is combined from multiple strips, and it involves details including camera specialties and image combination. It will be another project to complete.

(3) 3D visualization and interactive system

We have the 3D visualization tool for the depth map, but we need an integrated program to combine the stereo map tool and the visualization tool.

We believe that these two tools could be the base of the interactive system discussed in section VI. If the geoscientists could see the changes of the terrain immediately after they change the values in the depth map, it would be very useful for them to understand the terrain more.

(4) Parallel Computing and Divide and Conquer

There are some approaches to make our tool faster.

First is the parallel computing and the second is the "divide and conquer" technique.

We haven't used any parallel computing technique in our tool, and the bottleneck of the computational time in our program is the similarity values computation, which could be done by CPUs independently. Therefore, if we can use threads to compute the similarity values, we can save a lot of time.

Another way to accelerate our program is to divide the stereo pair images into pieces, and combine the results. However, we will need an elegant algorithm to merge the edges of the results so that the maximum-surface can be smooth. The blurring technique in image processing may be a effective way to do it.